Practical 3

1940223, Pranav G, 4 CMS

2021-01-29

popl = USArrests$Murder

# QUESTION 1

Draw a random sample of size 15 using with replacement procedure and estimate the average no. of murder arrests in the US and also obtain the 95% confidence limits for this.

sample = sample(popl, 6, replace = TRUE)

Average number of murders, based on this sample, is

sample\_mean = mean(sample)  
sample\_mean

## [1] 6.483333

To obtain the 95% confidence interval…

popl\_var = var(popl)  
popl\_size = length(popl)  
sample\_size = length(sample)  
popl\_mean\_square = ((popl\_size) / (popl\_size - 1)) \* popl\_var  
upper = sample\_mean + sqrt(((popl\_size - 1) / popl\_size) \* popl\_mean\_square / sample\_size) \* qt(0.05, sample\_size - 1, lower.tail = FALSE)  
lower = sample\_mean - sqrt(((popl\_size - 1) / popl\_size) \* popl\_mean\_square / sample\_size) \* qt(0.05, sample\_size - 1, lower.tail = FALSE)  
# We get the confidence interval as  
c(lower, upper)

## [1] 2.900317 10.066350

In other words, around 95% of our estimates of the average murder arrests in USA, inferred from samples of size 6, will be within the above interval.

# QUESTION 2

Show that the sample mean is unbiased for population mean in case of SRSWR.

An unbiased estimator is one whose expected value is equal to the parameter. To verify this, we will draw 1000,000 samples of size 50 from the population, and average their means.

sample\_means = replicate(1000000, mean(sample(popl, 50, replace = FALSE)))  
sample\_means\_mean = mean(sample\_means)  
popl\_mean = mean(popl)  
  
sample\_means\_mean

## [1] 7.788

popl\_mean

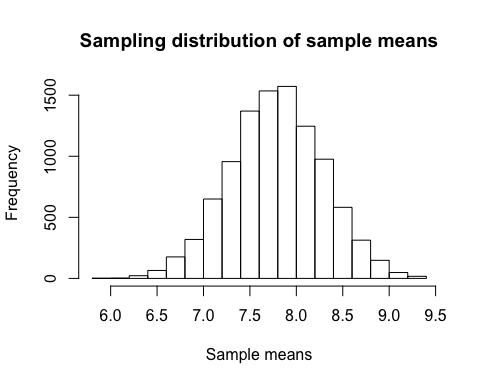
## [1] 7.788

Hence, we see that the average of the means of a million samples of size 50 is very close to the population mean. Hence, we may conclude that sample mean is an unbiased estimator of population mean in SRSWR.

# QUESTION 3

Obtain the sampling distribution of the estimate of average no. of murder arrests in the US and represent it by a histogram.

sample\_means = replicate(10000, mean(sample(popl, 30, replace = FALSE)))  
hist(main = "Sampling distribution of sample means", xlab = "Sample means", ylab = "Frequency", sample\_means)



# QUESTION 4

Verify that the SRSWOR provides the better estimate of population mean than the SRSWR (use formulae to calculate the variances in both cases, SRSWOR and SRSWR, by taking a sample of same size and compare the variances).

# Variance of means taken from 100 samples with replacement...  
vars\_mean\_srswr = var(replicate(100, mean(sample(popl, 10, replace = TRUE))))  
# Variance of means taken from 100 samples without replacement...  
vars\_mean\_srswor = var(replicate(100, mean(sample(popl, 10, replace = FALSE))))  
vars\_mean\_srswr

## [1] 2.074991

vars\_mean\_srswor

## [1] 1.326007

As we can see, the variance of means in SRSWOR is lesser than the variance of means in SRSWR. This indicates that the mean taken from samples without replacement is a more efficient estimator than the mean taken from samples with replacement.